

OICE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/873,737A

DATE: 10/23/2001

TIME: 10:49:29

Input Set : A:\180-104-2.ST25 revised 07-30-01.txt

Output Set: N:\CRF3\10232001\I873737A.raw

3 <110> APPLICANT: Duke University
 4 Lin, Haifan
 6 <120> TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
 7 PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
 9 <130> FILE REFERENCE: Attorney Docket No. 180-104/2
 11 <140> CURRENT APPLICATION NUMBER: 09/873,737A
 12 <141> CURRENT FILING DATE: 2001-06-04
 14 <150> PRIOR APPLICATION NUMBER: PCT/US99/28764
 15 <151> PRIOR FILING DATE: 1999-12-03
 17 <150> PRIOR APPLICATION NUMBER: 60/110,901
 18 <151> PRIOR FILING DATE: 1998-12-04
 20 <160> NUMBER OF SEQ ID NOS: 21
 22 <170> SOFTWARE: PatentIn Ver. 2.1
 25 <210> SEQ ID NO: 1
 26 <211> LENGTH: 3047
 27 <212> TYPE: DNA
 28 <213> ORGANISM: Drosophila sp.
 W--> 29 <220> FEATURE:
 30 <221> NAME/KEY: CDS
 31 <222> LOCATION: (84)..(2612)
 W--> 33 <220> FEATURE:
 34 <221> NAME/KEY: misc_feature
 35 <222> LOCATION: (120)
 36 <223> OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
 W--> 38 <220> FEATURE:
 39 <221> NAME/KEY: misc_feature
 40 <222> LOCATION: (399)
 41 <223> OTHER INFORMATION: n=a or t, Xaa=Leu or Ile
 W--> 43 <220> FEATURE:
 44 <221> NAME/KEY: misc_feature
 45 <222> LOCATION: (2436)
 46 <223> OTHER INFORMATION: n=a or c, Xaa=Leu or Ile
 48 <400> SEQUENCE: 1
 49 ctgagtccaa agcgtcggttt tcaaagtact ctttcagttt ccattgtgaa gttttaagt 60
 51 atcgcgagtgc ccaaaaagta aca atg gct gat gat cag gga cgt gga cgc agg 113
 52 Met Ala Asp Asp Gln Gly Arg Gly Arg Arg
 53 1 5 10
 W--> 55 cgt cca ntt aac gaa gat gat tcc tct act tcc cga ggt agt ggt gat 161
 W--> 56 Arg Pro Xaa Asn Glu Asp Asp Ser Ser Thr Ser Arg Gly Ser Gly Asp
 57 15 20 25
 59 ggg ccg cgg gtg aaa gta ttc aga gga tct tca tca ggt gac ccg aga 209
 60 Gly Pro Arg Val Lys Val Phe Arg Gly Ser Ser Ser Gly Asp Pro Arg
 61 30 35 40
 63 gcg gat cct cgt ata gag gct tca aga gag aga aga gct ctc gag gaa 257
 64 Ala Asp Pro Arg Ile Glu Ala Ser Arg Glu Arg Arg Ala Leu Glu Glu
 65 45 50 55
 67 gct ccc agg cgt gaa ggt ggc ccg cca gag cga aag ccg tgg ggt gac 305

ENTERED

p.5

RAW SEQUENCE LISTING

DATE: 10/23/2001

PATENT APPLICATION: US/09/873,737A

TIME: 10:49:29

Input Set : A:\180-104-2.ST25 revised 07-30-01.txt

Output Set: N:\CRF3\10232001\I873737A.raw

```

68 Ala Pro Arg Arg Glu Gly Gly Pro Pro Glu Arg Lys Pro Trp Gly Asp
69      60      65      70
71 caa tat gat tac ctg aat acc cgt ccg gtt gag ctg gta tcc aag aag 353
72 Gln Tyr Asp Tyr Leu Asn Thr Arg Pro Val Glu Leu Val Ser Lys Lys
73 75      80      85
W--> 75 gga acc gat ggc gtc ccg gtc atg ctg cag acg aac ttt ttt cga nta 401
W--> 76 Gly Thr Asp Gly Val Pro Val Met Leu Gln Thr Asn Phe Phe Arg Xaa
77      95      100      105
79 aaa acc aag ccg gaa tgg cgg atc gtt cat tat cac gtg gag ttt gtg 449
80 Lys Thr Lys Pro Glu Trp Arg Ile Val His Tyr His Val Glu Phe Val
81      110      115      120
83 ccg acc atc gag aat cct cgt gtc cgt atg gga gtt ttg tcc aat cat 497
84 Pro Thr Ile Glu Asn Pro Arg Val Arg Met Gly Val Leu Ser Asn His
85      125      130      135
87 gct aac ctt ctg gga tca ggc tat cta ttc gac gga ctg caa ctg ttc 545
88 Ala Asn Leu Leu Gly Ser Gly Tyr Leu Phe Asp Gly Leu Gln Leu Phe
89      140      145      150
91 acc acc agg aaa ttc gag cag gaa atc acg gtg ctc agc gga aag tcg 593
92 Thr Thr Arg Lys Phe Glu Gln Glu Ile Thr Val Leu Ser Gly Lys Ser
93 155      160      165      170
95 aag ctg gac att gaa tac aag ata tcc ata aag ttc gtt gga ttc ata 641
96 Lys Leu Asp Ile Glu Tyr Lys Ile Ser Ile Lys Phe Val Gly Phe Ile
97      175      180      185
99 tcg tgt gct gag ccc cgc ttt ttg caa gtc tta aat cta ata ttg cgc 689
100 Ser Cys Ala Glu Pro Arg Phe Leu Gln Val Leu Asn Leu Ile Leu Arg
101      190      195      200
103 cgc tcg atg aag ggc cta aat ttg gaa tta gtt ggc cgt aat ctc ttt 737
104 Arg Ser Met Lys Gly Leu Asn Leu Glu Leu Val Gly Arg Asn Leu Phe
105      205      210      215
107 gat ccc cga gct aag atc gaa ata agg gag ttc aaa atg gag cta tgg 785
108 Asp Pro Arg Ala Lys Ile Glu Ile Arg Glu Phe Lys Met Glu Leu Trp
109      220      225      230
111 ccg ggc tat gag aca tcg att cgt cag cac gaa aaa gat att tta ttg 833
112 Pro Gly Tyr Glu Thr Ser Ile Arg Gln His Glu Lys Asp Ile Leu Leu
113 235      240      245      250
115 ggc acc gaa ata act cac aaa gtt atg cgc acc gag acg atc tac gac 881
116 Gly Thr Glu Ile Thr His Lys Val Met Arg Thr Glu Thr Ile Tyr Asp
117      255      260      265
119 ata atg cga cgt tgc tca cac aat ccg gct cgt cat cag gac gaa gta 929
120 Ile Met Arg Arg Cys Ser His Asn Pro Ala Arg His Gln Asp Glu Val
121      270      275      280
123 cgg gta aat gtt ttg gac ttg att gtc ctt acg gat tac aat aac aga 977
124 Arg Val Asn Val Leu Asp Leu Ile Val Leu Thr Asp Tyr Asn Asn Arg
125      285      290      295
127 act tat cgt atc aat gat gtc gac ttt gga caa act ccg aaa tca aca 1025
128 Thr Tyr Arg Ile Asn Asp Val Asp Phe Gly Gln Thr Pro Lys Ser Thr
129      300      305      310
131 ttc agt tgc aag ggt aga gat atc agt ttc gtg gaa tac tat ctc act 1073
132 Phe Ser Cys Lys Gly Arg Asp Ile Ser Phe Val Glu Tyr Tyr Leu Thr

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/873,737A

DATE: 10/23/2001

TIME: 10:49:29

Input Set : A:\180-104-2.ST25 revised 07-30-01.txt

Output Set: N:\CRF3\10232001\I873737A.raw

133	315		320		325		330	
135	aaa	tat	aat	ata	cgc	att	cgc	gac
136	Lys	Tyr	Asn	Ile	Arg	Ile	Arg	Asp
137								
139	aaa	aat	agg	gac	aag	gct	cta	aaa
140	Lys	Asn	Arg	Asp	Lys	Ala	Leu	Lys
141								
143	cta	att	cct	gag	ctc	tgc	cga	gtg
144	Leu	Ile	Pro	Glu	Leu	Cys	Arg	Val
145								
147	tca	aac	ttt	cag	ctt	atg	cgt	gcc
148	Ser	Asn	Phe	Gln	Leu	Met	Arg	Ala
149								
151	ccc	aaa	caa	cgc	act	gat	cga	ttg
152	Pro	Lys	Gln	Arg	Thr	Asp	Arg	Leu
153	395							
155	aac	act	cca	gaa	agt	gtg	aag	gtc
156	Asn	Thr	Pro	Glu	Ser	Val	Lys	Val
157								
159	gac	aag	aac	gtc	aca	gaa	gta	caa
160	Asp	Lys	Asn	Val	Thr	Glu	Val	Gln
161								
163	atc	gtg	ttt	cat	aat	gga	aag	gtt
164	Ile	Val	Phe	His	Asn	Gly	Lys	Val
165								
167	caa	agg	cac	ttc	aga	gac	caa	agg
168	Gln	Arg	His	Phe	Arg	Asp	Gln	Arg
169								
171	ctc	gat	cgt	tgg	gct	gtc	atc	gcg
172	Leu	Asp	Arg	Trp	Ala	Val	Ile	Ala
173	475							
175	cga	act	cta	ctt	gac	tct	ttg	tat
176	Arg	Thr	Leu	Leu	Asp	Ser	Leu	Tyr
177								
179	aga	att	cga	agc	ccc	cag	gaa	ttc
180	Arg	Ile	Arg	Ser	Pro	Gln	Glu	Phe
181								
183	act	tat	gtg	aga	gca	atg	gat	gat
184	Thr	Tyr	Val	Arg	Ala	Met	Asp	Asp
185								
187	ata	tta	tgc	ctc	gta	ccc	aat	gat
188	Ile	Leu	Cys	Leu	Val	Pro	Asn	Asp
189								
191	aaa	aag	aga	gga	tac	gtt	gac	agg
192	Lys	Lys	Arg	Gly	Tyr	Val	Asp	Arg
193	555							
195	ctt	aaa	acg	acc	aag	aac	cgt	agc
196	Leu	Lys	Thr	Thr	Lys	Asn	Arg	Ser
197								

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/873,737A

DATE: 10/23/2001

TIME: 10:49:29

Input Set : A:\180-104-2.ST25 revised 07-30-01.txt

Output Set: N:\CRF3\10232001\I873737A.raw

```

199 gca atc caa ctg aat tgc aag ttg gga tat aca ccc tgg atg atc gaa 1889
200 Ala Ile Gln Leu Asn Cys Lys Leu Gly Tyr Thr Pro Trp Met Ile Glu
201          590          595          600
203 cta ccc ttg tcc gga ctg atg aca att ggc ttt gac att gcg aag agc 1937
204 Leu Pro Leu Ser Gly Leu Met Thr Ile Gly Phe Asp Ile Ala Lys Ser
205          605          610          615
207 aca cga gat cgg aag agg gcc tac gga gca ttg att gcc tca atg gat 1985
208 Thr Arg Asp Arg Lys Arg Ala Tyr Gly Ala Leu Ile Ala Ser Met Asp
209          620          625          630
211 cta cag caa aac tcc acg tac ttc agc aca gtc acg gag tgc agc gcc 2033
212 Leu Gln Gln Asn Ser Thr Tyr Phe Ser Thr Val Thr Glu Cys Ser Ala
213 635          640          645          650
215 ttt gat gtg ctc gct aac acc ctt tgg ccg atg ata gca aag gcc ctg 2081
216 Phe Asp Val Leu Ala Asn Thr Leu Trp Pro Met Ile Ala Lys Ala Leu
217          655          660          665
219 cgc caa tat caa cat gag cat agg aag ctg cca tct cga atc gta ttt 2129
220 Arg Gln Tyr Gln His Glu His Arg Lys Leu Pro Ser Arg Ile Val Phe
221          670          675          680
223 tat cga gac ggt gtg agc tcc ggc tct cta aag cag ctt ttt gaa ttt 2177
224 Tyr Arg Asp Gly Val Ser Ser Gly Ser Leu Lys Gln Leu Phe Glu Phe
225          685          690          695
227 gaa gtc aag gac atc att gag aag ttg aaa act gaa tac gcc cgc gtc 2225
228 Glu Val Lys Asp Ile Ile Glu Lys Leu Lys Thr Glu Tyr Ala Arg Val
229          700          705          710
231 cag cta agc cca ccg caa tta gct tat att gtg gta acc aga tcc atg 2273
232 Gln Leu Ser Pro Pro Gln Leu Ala Tyr Ile Val Val Thr Arg Ser Met
233 715          720          725          730
235 aac acg cgc ttc ttc ctc aac gga caa aat cct ccg cct ggt act ata 2321
236 Asn Thr Arg Phe Phe Leu Asn Gly Gln Asn Pro Pro Pro Gly Thr Ile
237          735          740          745
239 gtt gat gac gtt ata act ctg ccc gag aga tac gac ttt tat ctg gtc 2369
240 Val Asp Asp Val Ile Thr Leu Pro Glu Arg Tyr Asp Phe Tyr Leu Val
241          750          755          760
243 tcg caa caa gtt cgt cag ggt aca gtg tcg ccg acc agc tac aat gtt 2417
244 Ser Gln Gln Val Arg Gln Gly Thr Val Ser Pro Thr Ser Tyr Asn Val
245          765          770          775
W--> 247 ctt tat agc agc atg ggt ntc tca ccg gag aaa atg caa aaa ctt acg 2465
W--> 248 Leu Tyr Ser Ser Met Gly xaa Ser Pro Glu Lys Met Gln Lys Leu Thr
249          780          785          790
251 tac aag atg tgc cac ttg tac tac aat tgg tcg ggc acc aca cga gtg 2513
252 Tyr Lys Met Cys His Leu Tyr Tyr Asn Trp Ser Gly Thr Thr Arg Val
253 795          800          805          810
255 cca gca gtt tgc cag tac gct aag aag cta gct acc ctc gtg ggt acg 2561
256 Pro Ala Val Cys Gln Tyr Ala Lys Lys Leu Ala Thr Leu Val Gly Thr
257          815          820          825
259 aac ttg cac tct att ccg caa aac gcg ctc gaa aag aag ttt tat tat 2609
260 Asn Leu His Ser Ile Pro Gln Asn Ala Leu Glu Lys Lys Phe Tyr Tyr
261          830          835          840
263 cta taattggata taatttagaa tggagtatta atccttacta agaggccata 2662

```

RAW SEQUENCE LISTING

DATE: 10/23/2001

PATENT APPLICATION: US/09/873,737A

TIME: 10:49:29

Input Set : A:\180-104-2.ST25 revised 07-30-01.txt

Output Set: N:\CRF3\10232001\I873737A.raw

264 Leu
 266 tatgaaacta gccagacat ttatactttt tcaatacttc cttacttttg ctaagcactt 2722
 268 cagcatttat gactaaatat tttgtatttg aaatgcatta ctgctctttt ttcaaacaaa 2782
 270 agcaaaattg aggattaaga ttctggtatt taagcataag accagaggaa attcccaaac 2842
 272 aaacatttaa agttatctat caagacatgt tcattaattt ggaatataat tactttattt 2902
 274 tttattgtat attttagttt atgtaaagaa aaattacata catccatggt tgcttactta 2962
 276 accacacatt catggctgct tatattcgtg aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3022
 278 aaaaaaaaaa aaaaaaaaaa aaaaaa 3047

281 <210> SEQ ID NO: 2

282 <211> LENGTH: 843

283 <212> TYPE: PRT

284 <213> ORGANISM: Drosophila sp.

286 <220> FEATURE:

287 <221> NAME/KEY: misc_feature

288 <222> LOCATION: (13)

289 <223> OTHER INFORMATION: Xaa=Leu or Ile

291 <220> FEATURE:

292 <221> NAME/KEY: misc_feature

293 <222> LOCATION: (106)

294 <223> OTHER INFORMATION: Xaa=Leu or Ile

296 <220> FEATURE:

297 <221> NAME/KEY: misc_feature

298 <222> LOCATION: (785)

299 <223> OTHER INFORMATION: Xaa=Leu or Ile

301 <400> SEQUENCE: 2

W--> 302 Met Ala Asp Asp Gln Gly Arg Gly Arg Arg Arg Pro Xaa Asn Glu Asp
 303 1 5 10 15
 305 Asp Ser Ser Thr Ser Arg Gly Ser Gly Asp Gly Pro Arg Val Lys Val
 306 20 25 30
 308 Phe Arg Gly Ser Ser Ser Gly Asp Pro Arg Ala Asp Pro Arg Ile Glu
 309 35 40 45
 311 Ala Ser Arg Glu Arg Arg Ala Leu Glu Glu Ala Pro Arg Arg Glu Gly
 312 50 55 60
 314 Gly Pro Pro Glu Arg Lys Pro Trp Gly Asp Gln Tyr Asp Tyr Leu Asn
 315 65 70 75 80
 317 Thr Arg Pro Val Glu Leu Val Ser Lys Lys Gly Thr Asp Gly Val Pro
 318 85 90 95
 W--> 320 Val Met Leu Gln Thr Asn Phe Phe Arg Xaa Lys Thr Lys Pro Glu Trp
 321 100 105 110
 323 Arg Ile Val His Tyr His Val Glu Phe Val Pro Thr Ile Glu Asn Pro
 324 115 120 125
 326 Arg Val Arg Met Gly Val Leu Ser Asn His Ala Asn Leu Leu Gly Ser
 327 130 135 140
 329 Gly Tyr Leu Phe Asp Gly Leu Gln Leu Phe Thr Thr Arg Lys Phe Glu
 330 145 150 155 160
 332 Gln Glu Ile Thr Val Leu Ser Gly Lys Ser Lys Leu Asp Ile Glu Tyr
 333 165 170 175
 335 Lys Ile Ser Ile Lys Phe Val Gly Phe Ile Ser Cys Ala Glu Pro Arg
 336 180 185 190

Use of n and / or Xaa has been detected in the
 Sequence Listing. Review the Sequence Listing
 to ensure a corresponding explanation is present
 in the <220> to <223> fields of each sequence
 using n or Xaa.

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/873,737A

DATE: 10/23/2001

TIME: 10:49:30

Input Set : A:\180-104-2.ST25 revised 07-30-01.txt

Output Set: N:\CRF3\10232001\I873737A.raw

L:29 M:283 W: Missing Blank Line separator, <220> field identifier
L:55 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:56 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:75 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:76 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:247 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:248 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1
L:302 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:320 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:449 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2
L:466 M:283 W: Missing Blank Line separator, <220> field identifier
L:517 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:518 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:549 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:550 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:593 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:594 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:701 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:702 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3
L:800 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4
L:824 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4
L:854 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4
L:935 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:4
L:952 M:283 W: Missing Blank Line separator, <220> field identifier
L:996 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:997 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:1052 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:1053 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:1160 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:1161 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:5
L:1254 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6
L:1296 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6
L:1377 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:6
L:1409 M:283 W: Missing Blank Line separator, <220> field identifier
L:1421 M:283 W: Missing Blank Line separator, <220> field identifier
L:1433 M:283 W: Missing Blank Line separator, <220> field identifier
L:1445 M:283 W: Missing Blank Line separator, <220> field identifier
L:1457 M:283 W: Missing Blank Line separator, <220> field identifier
L:1469 M:283 W: Missing Blank Line separator, <220> field identifier
L:1481 M:283 W: Missing Blank Line separator, <220> field identifier
L:1493 M:283 W: Missing Blank Line separator, <220> field identifier
L:1505 M:283 W: Missing Blank Line separator, <220> field identifier
L:1517 M:283 W: Missing Blank Line separator, <220> field identifier
L:1529 M:283 W: Missing Blank Line separator, <220> field identifier
L:1541 M:283 W: Missing Blank Line separator, <220> field identifier
L:1553 M:283 W: Missing Blank Line separator, <220> field identifier
L:1565 M:283 W: Missing Blank Line separator, <220> field identifier
L:1576 M:283 W: Missing Blank Line separator, <220> field identifier